http://enes.org/

Copernicus and H2020 Program: Machine Learning and Big Data Needs and Overview

S.Denvil[1], M.Lautenschlager[2], S.Fiore[4], F.Guglielmo[1], S.Joussaume[1], M.Juckes[2], S.Kinderman[2], M.Kolax[7], C.Pagé[3], W.Som de Cerff[6].

[1]Institut Pierre Simon Laplace, IPSL, France
[2]Centre for Environmental Data Analysis, CEDA, United Kingdom
[3]Centre Européen de recherche et de formation avancée en calcul scientifique, CERFACS, France
[4]Euro-Mediterranean Center on Climate Change Foundation, CMCC, Italy
[5]Deutsches Klimarechenzentrum, DKRZ, Germany
[6]The Royal Netherlands Meteorological Institute, KNMI, Netherland
[7]Swedish Meteorological and Hydrological Institute, SMHI, Sweden

New topic for exploration for IS-ENES3

The use of machine learning in climate modelling is highly innovative and work in this area is in its infancy.

However, machine learning as a concept is well established.

To our knowledge, machine learning is not currently used routinely by any major modelling centre, however the benefits might well be considerable.

Feed Forward Neural Network Model

LSCE/CNRS/IPSL: Anna Sommer, Marion Gehlen, Mathieu Vrac, Carlos Mejia

- Global Reconstruction of pCO_2
- Identification of an optimal observational network for enabling ocean carbon system estimates
- Period at the moment: 2001-2016

FFNN: pCO_2 anomalies as non-linear function of oceanic and atmospheric drivers

$$pCO_{2\,Anom} =$$

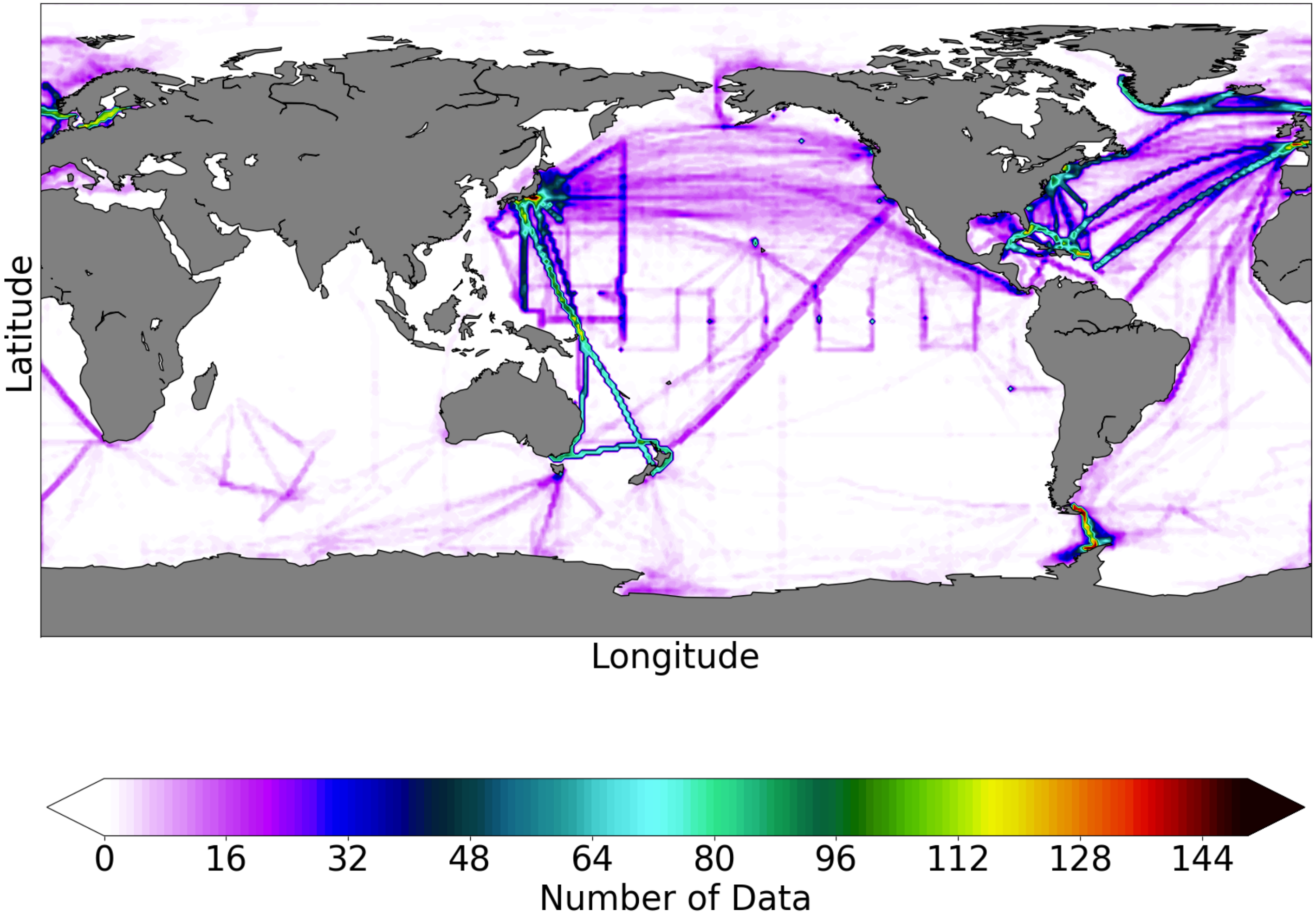
$$g(SSS, SST, SSH, MLD, Chl, CO2, Atm, lon, lat, SSS_{Anom}, SST_{Anom}, SSH_{Anom}, MLD_{Anom}, Chl_{Anom}, CO_{2,Atm\ Anom})$$

Target pCO2, ocean

Observation data: SOCAT v5 – ship traces

for period Jan 1970-Dec 2016

Chosen period: 2001-2016 (25 % of data for validation)

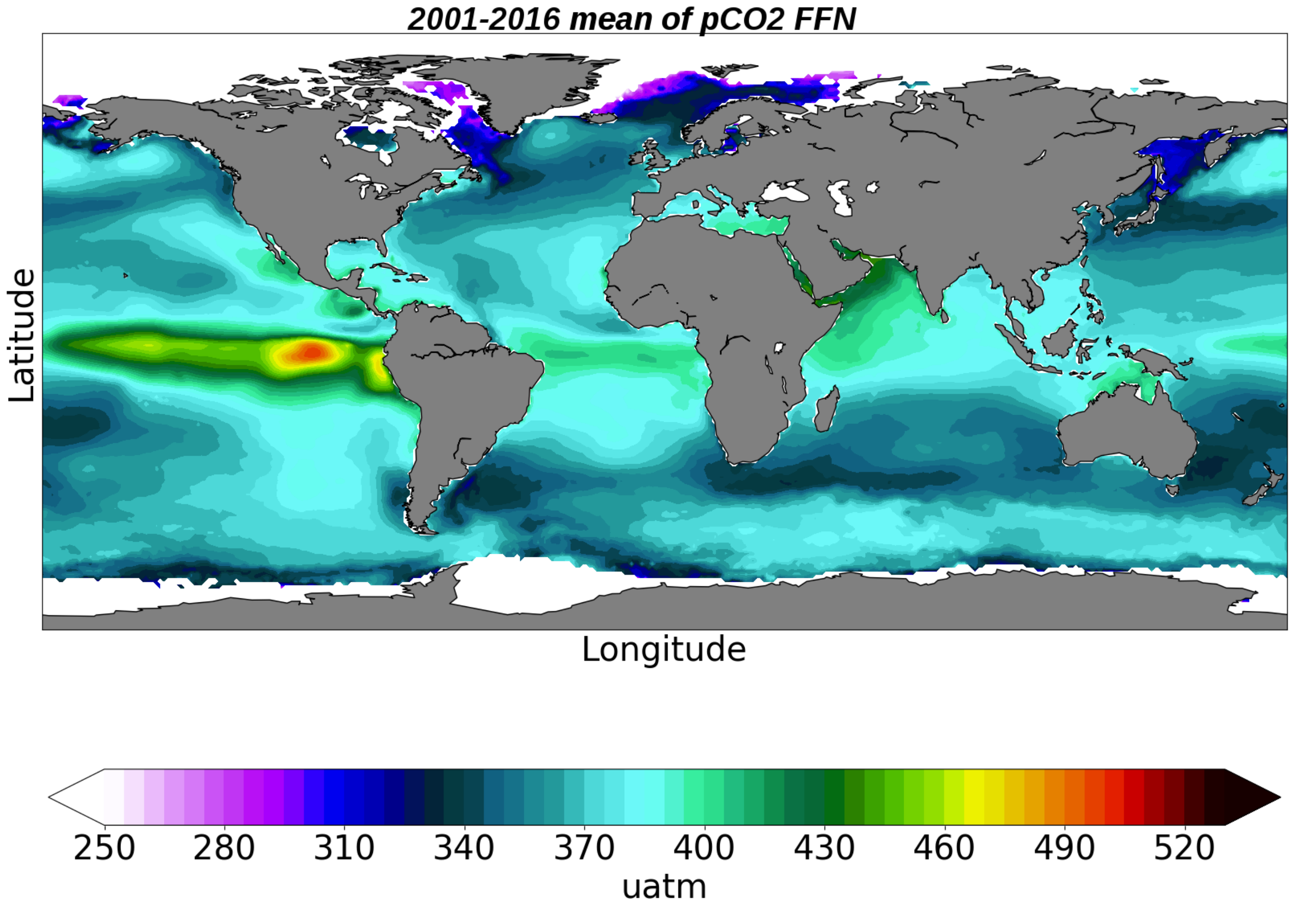


FFNN pCO2,ocean

RMS = 18.15 uatm

R^2 = 0.75

Bias = 0.99



Cloud computing & Big Data

Domains of interest to weather and climate modeling community:

New ways of exploiting algorithms **emerging** (in our community).

- Using machine learning to identifying patterns in data, something we've done for decades, but with new and (possibly) better tools.
- Using machine learning for Quality Control of data (e.g. unusual field)
- Using machine learning to improve efficiency of physics code
- Using machine learning to inform climate model development

Improving efficiency of physics code

Emulating the physics of existing general circulation models (GCMs) to increase computational efficiency of the code when running ‘operational’ simulations.

Conduct simulations with more ensemble members or at higher resolution for the same computational cost.

Investigate the most appropriate method to train the emulator (single column model simulations, full GCM from an initialised state, etc.).

Investigate how to link existing machine learning libraries into operational simulations

Run a full climate simulation using emulated physics with success criteria being the code running faster and scientific results not unacceptably degraded.

Inform climate model development

Machine learning offers a whole class of new statistical tools for doing model/observation comparison on large datasets.

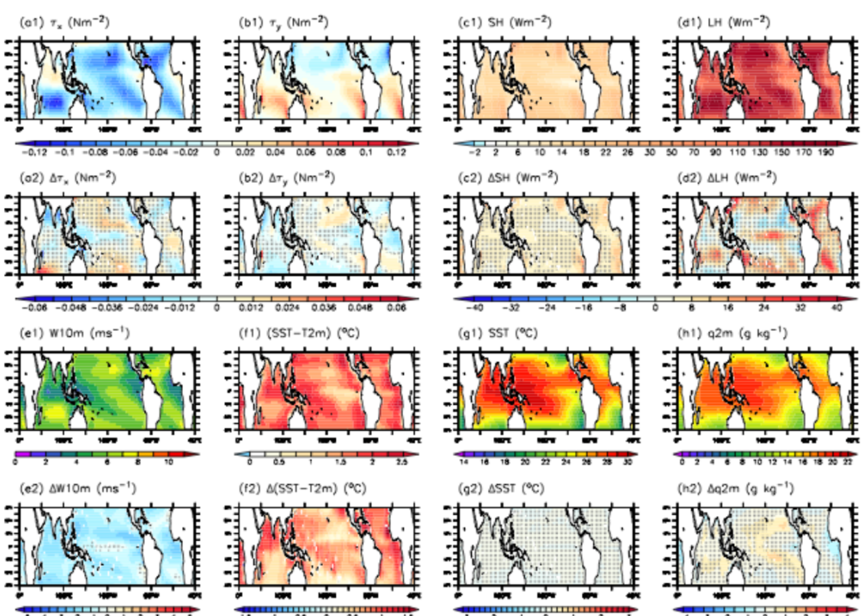
Another way of evaluating model output is on the level of causal interactions.

A hot topic in machine learning is causal inference which provides methods to extract the causal interdependency structure of the data.

Typically, such methods would be applied to dimensionally reduced model output and reanalysis/observational data.

Model evaluation can then be based on comparing the causal interdependencies within the reanalysis/observations with the causal interdependencies in the model output.

Questions ?


(a1) τ_x (Nm^{-2})
(b1) τ_y (Nm^{-2})
(c1) SH (Wm^{-2})
(d1) LH (Wm^{-2})
(a2) $\Delta\tau_x$ (Nm^{-2})
(b2) $\Delta\tau_y$ (Nm^{-2})
(c2) ΔSH (Wm^{-2})
(d2) ΔLH (Wm^{-2})
(e1) W10m (ms^{-1})
(f1) (SST−T2m) (ºC)
(g1) SST (ºC)
(h1) q2m (g kg^{-1})
(e2) ΔW10m (ms^{-1})
(f2) Δ(SST−T2m) (ºC)
(g2) ΔSST (ºC)
(h2) Δq2m (g kg^{-1})